Extending functional genome annotations using high-throughput CRISPRi

Harneet S. Rishi[1,2*] (HSRishi@lbl.gov), Esteban Toro,[3] Honglei Liu,[4] Xiaowo Wang,[4,5] Lei S. Qi,[6,7] and **Adam P. Arkin**[2,3]

[1]Biophysics Graduate Program, University of California Berkeley, Berkeley, CA; [2]Environmental Genomics and Systems Biology Division, Lawrence Berkeley National Laboratory, Berkeley, CA; [3]Department of Bioengineering, University of California Berkeley, Berkeley, CA; [4]Bioinformatics Division, Center for Synthetic and Systems Biology, Tsinghua National Laboratory for Information Science and Technology, Beijing, China; [5]Department of Automation, Tsinghua University, Beijing, China; [6]Department of Cellular and Molecular Pharmacology, University of California San Francisco, San Francisco, CA; and [7]California Institute for Quantitative Biomedical Research, San Francisco, CA

http://genomicscience.energy.gov/biosystemsdesign/2012awards.shtml

Project Goals: Synthetic biology aims to leverage the engineering principles of modularity, standardization, and reliability with the design-build-test-learn cycle to rapidly engineer novel biological functions. One of the key hurdles in adopting this strategy is that the innate complexity of biological systems (ex: unmapped interaction networks, context dependence, temporal & spatial population variations) makes it difficult to understand first principles, which in turn makes it difficult to predictably build scalable systems. High-throughput technologies to quantitatively characterize sequence-function-phenotype landscapes can help overcome such barriers. As such, we aim to develop CRISPR interference (CRISPRi) as a platform for pooled bacterial functional genomics with the goal of investigating combinatorial genetic interactions in high-throughput.

The advent of next-generation sequencing has led to an explosion of genome sequences, and now reverse-genetics efforts are becoming increasingly vital to finding the phenotype and, by extension, function associated with a given gene of interest. As a first pass, computational workflows can predict the location of genes and regulatory features in a genome, and automated annotation pipelines can associate hypothetical functions with a given gene based on sequence homology. However, such *in silico* predictions can sometimes be erroneous – necessitating the experimental validation of predicted gene functions. Large mutant collections, which often comprise of gene disruptions in the form of knockouts or insertions, can complement computational prediction pipelines by affording the ability to experimentally interrogate many genetic perturbations in a native context.

The compactness, modularity, and largely species-independent functionality (transcriptional regulation, genome editing, imaging) of the CRISPR system make it an attractive tool for genome-wide screens. Here, we apply the catalytically inactive dCas9 to conduct high-throughput transcriptional and regulatory studies in *E. coli*. Using an Agilent OLS library of 32992 unique sgRNAs, we targeted 4500 genes, 5400 promoters, 640 transcription factor

binding sites (TFBSs), and 106 small RNAs (sRNAs) in the *E. coli* genome. These genomic targets cover a wide range of functionalities such as metabolism, stress response, transport, and cell division. By combining CRISPRi with next-generation sequencing, we were able to interrogate the fitness effect of transcriptional knockdown for each of the aforementioned genomic features in a single-pot experiment both aerobically and anaerobically. Our fitness results agreed well with current knockout databases, and our ability to induce transcriptional knockdown at any point during an experiment has allowed us to explore target essentiality under different conditions with great ease. We demonstrate this by showing that although *nrdA* and *nrdB* are essential under aerobic conditions – and are annotated as such in databases – they are dispensable anaerobically. We also show that CRISPRi can recapitulate genomic features with redundant functions, thus demonstrating our ability to interrogate combinatorial interactions. Additionally, we explore the efficacy of non-genic CRISPRi on a genome-wide scale and create cofitness profiles linking together promoters, TFBSs, and genes in transcriptional units.

We also leverage the CRISPRi library with flow cytometry to obtain non-growth phenotypes such as morphology in high-throughput. We first show that CRISPRi can generate filamentous cellular phenotypes and next present a flow-seq methodology that allows us to enrich for filamentous mutants in our library. Finally we show that our single-pot flow-seq results agree well with single-genotype microscopy studies and provide novel phenotypes for several essential genes and genes of unknown function.

Overall, HT-CRISPRi enables single-pot, precise measurements of fitness for a large set of genomic features and will prove useful in genomic studies of model and non-model organisms. The extension of the platform to perform genetic interactions should also provide a comprehensive testbed for exploring epistatic landscapes and revealing the basis of complex traits.

This work is supported by the Genome Science program within the Office of Biological and Environmental Research (Project grant number DE-SC008812, Funding Opportunity Announcement DE-FOA-0000640*).*